On the Effectiveness of Projection Methods for Convex Feasibility Problems with Linear Inequality Constraints

Y. Censor,[1] W. Chen,[2] P. L. Combettes,[3] R. Davidi,[2] and G. T. Herman[2]

[1]Department of Mathematics, University of Haifa, Mt. Carmel Haifa 31905, Israel
[2]Department of Computer Science, The Graduate Center, City University of New York
365 Fifth Avenue, New York, NY 10016, USA
[3]UPMC Université Paris 06, Laboratoire Jacques-Louis Lions – UMR CNRS 7598
75005 Paris, France

22 décembre 2009

Résumé

The effectiveness of projection methods for solving systems of linear inequalities is investigated. It is shown that they have a computational advantage over some alternatives and that this makes them successful in real-world applications. This is supported by experimental evidence provided in this paper on problems of various sizes (up to tens of thousands of unknowns satisfying up to hundreds of thousands of constraints) and by a discussion of the demonstrated efficacy of projection methods in numerous scientific publications and commercial patents (dealing with problems that can have over a billion unknowns and a similar number of constraints).

1 Introduction

Projection methods were first used to solve systems of linear equations in Euclidean spaces in the 1930s [31, 55] and were subsequently extended to systems of linear inequalities in [1, 62, 63]. The basic step in these early algorithms consists of a projection onto an affine subspace or a half-space. Modern projection methods are much more sophisticated [7, 8, 9, 10, 17, 26, 34, 35, 36, 42, 43, 57] and they can solve the general *convex feasibility problem* of finding a point in the intersection of a family of closed convex sets in a Hilbert space. In such formulations, each set can be specified in various forms, e.g., as the fixed point set of a nonexpansive operator, the set of zeros of a maximal monotone operator, the set of solutions to a convex inequality, or the set of solutions to an equilibrium problem. Projection methods can have various algorithmic structures (some of which are particularly suitable for parallel computing) and they also possess desirable convergence properties and good initial behavior patterns [8, 26, 33, 34, 35, 51, 67]. The main advantage of projection methods, which makes them successful in real-world applications, is computational. They commonly have the ability to handle huge-size problems of dimensions beyond which more sophisticated methods cease to be efficient or even applicable due to memory requirements. This is so because the building bricks of a projection algorithm are the projections onto the given individual sets, which are assumed to be easy to perform, and because the algorithmic structure is either sequential or simultaneous, or in-between, as in the block-iterative

projection methods or in the more recently invented string-averaging projection methods. The number of sets used simultaneously in each iteration in block-iterative methods and the number and lengths of strings used in each iteration in string-averaging methods are variable, which provides great flexibility in matching the implementation of the algorithm with the parallel architecture at hand; for block-iterative methods see, e.g., [2, 10, 16, 23, 35, 40, 44, 46, 57, 64] and for string-averaging methods see, e.g., [12, 18, 22, 24, 25, 39, 65, 68].

The convex feasibility formalism is at the core of the modeling of many problems in various areas of mathematics and the physical sciences; see [32, 33] and references therein. Over the past four decades, it has been used to model significant real-world problems in sensor networks [14], in radiation therapy treatment planning [21, 52], in resolution enhancement [27], in wavelet-based denoising [30], in antenna design [49], in computerized tomography [51], in materials science [56], in watermarking [58], in data compression [60], in demosaicking [61], in magnetic resonance imaging [69], in holography [70], in color imaging [71], in optics and neural networks [72], in graph matching [73] and in adaptive filtering [75], to name but a few. In these – and numerous other – problems, projection methods have been used to solve the underlying convex feasibility problems.

We focus on the important subclass of convex feasibility problems in which finitely many sets are given and each of them is specified by a linear equality or inequality in the Euclidean space $\mathbb{R}^N$. For such problems, which arise in many important applications [32, 51, 52], alternatives to projection methods are available (see, e.g., [3, 48] and the references therein), and it is therefore legitimate to ask whether projection methods are competitive.

In this paper we address this question and show that projection methods are indeed very competitive in the environment of linear inequality constraints. In Section 2 we discuss their comparative performance for four different kinds of problems. In Section 3 we give some examples of their use in real-world applications from the research and the patent literature. Finally, we present our conclusions.

2 Comparisons

2.1 Examples of 2-set feasibility problems

In a recent paper [48], the author asks in the title : "How good are projection methods for convex feasibility problems ?" and immediately (in the Abstract) states that :

> "Unfortunately, particularly given the large literature which might make one think otherwise, numerical tests indicate that in general none of the variants [of projection methods for solving convex feasibility problems] considered are especially effective or competitive with more sophisticated alternatives."

As indicated in the Introduction, projection methods have been used to solve highly nonlinear complex problems involving a very large number of sets. Therefore, results based on the geometrically simple 2-set problems of [48] are vastly insufficient to draw general conclusions. In addition, we show in this subsection that the experiments reported in [48] use suboptimal versions of projection methods, which further questions the justification of the above-quoted general conclusion as to their effectiveness.

The numerical experiments provided in [48] focus exclusively on the problem of solving a linear system of equations under a box constraint, namely

$$\text{find } x \in \mathbb{R}^N, \text{ such that } \begin{cases} Ax = b, \\ x \in \bigtimes_{i=1}^{N} [c_i, d_i], \end{cases} \tag{1}$$

where $A \in \mathbb{R}^{M\times N}$ $(M \leq N)$ has full rank, $b \in \mathbb{R}^M$, and the problem is assumed to be feasible. We show that, even in this basic setting, projection algorithms implemented with standard relaxation strategies perform much better than indicated by the results in [48].

Let us denote by P_1 and P_2 the projection operators onto the closed affine subspace $S_1 = \left\{x \in \mathbb{R}^N \;\middle|\; Ax = b\right\}$ and the closed convex set $S_2 = \bigtimes_{i=1}^N [c_i, d_i]$, respectively. The first operator is defined by

$$P_1 \colon x \mapsto x - A^\top \left(AA^\top\right)^{-1}(Ax - b), \tag{2}$$

where $A^\top$ denotes the transpose of A. This transformation can be implemented in various fashions. For instance, in many signal and image processing problems, the matrix A is block-circulant and hence diagonalized by the discrete Fourier transform operator, which leads to a very efficient implementation of P_1 [4]. Here, we adopt a QR decomposition approach. Let

$$A^\top = \begin{bmatrix} Q_{11} & Q_{12} \end{bmatrix} \begin{bmatrix} R_{11} \\ 0 \end{bmatrix} \tag{3}$$

be the QR decomposition of $A^\top$, where R_{11} is an $M \times M$ invertible upper triangular matrix [45]. Then (2) yields

$$P_1 \colon x \mapsto x - Q_{11} \left(R_{11}^\top\right)^{-1} (Ax - b). \tag{4}$$

On the other hand, the projection $P_2 x = (\pi_i)_{1\leq i\leq N}$ of a vector $x = (x_i)_{1\leq i\leq N}$ onto S_2 is obtained through a simple clipping of its components, i.e., for every $i \in \{1, \ldots, N\}$, $\pi_i = \min\{\max\{x_i, c_i\}, d_i\}$.

Two standard projection methods to solve (1) are the alternating projection method

$$x^{(0)} \in \mathbb{R}^N \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x^{(n+1)} = x^{(n)} + \lambda_n (P_1 P_2 x^{(n)} - x^{(n)}) \tag{5}$$

and the parallel projection method

$$x^{(0)} \in \mathbb{R}^N \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x^{(n+1)} = x^{(n)} + \lambda_n \left(\frac{P_1 x^{(n)} + P_2 x^{(n)}}{2} - x^{(n)} \right), \tag{6}$$

where $(\lambda_n)_{n\in\mathbb{N}}$ is a sequence of strictly positive relaxation parameters. If $\lambda_n \equiv 1$ in (5), we obtain the popular *Projection Onto Convex Sets* (POCS) algorithm [32, 74] :

$$x^{(0)} \in \mathbb{R}^N \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x^{(n+1)} = P_1 P_2 x^{(n)}. \tag{7}$$

The convergence of any sequence $(x^{(n)})_{n\in\mathbb{N}}$ thus constructed to a point in $S_1 \cap S_2$ was established in [15]. On the other hand, if $\lambda_n \equiv 1$ in (6), we obtain the *Parallel Projection Method* (PPM) :

$$x^{(0)} \in \mathbb{R}^N \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x^{(n+1)} = \frac{P_1 x^{(n)} + P_2 x^{(n)}}{2}. \tag{8}$$

The convergence of any sequence $(x^{(n)})_{n\in\mathbb{N}}$ thus constructed to a point in $S_1 \cap S_2$ was established in [5], see also [6]. In [48], (5) and (6) are used, together with variants featuring a construction of λ_n at iteration n resulting from a line search procedure and without closed-form expression. However, as the numerical results of [48] show, these relaxation schemes do not lead to significantly better convergence profiles than those obtained with the unrelaxed algorithms POCS (7) and PPM (8). In addition, nothing is said regarding the convergence of (5) and (6) with such relaxation schemes.

The potentially slow convergence of projections methods has long been recognized [38, 50, 62] and remedies have been proposed to address this problem in the form of adapted relaxation

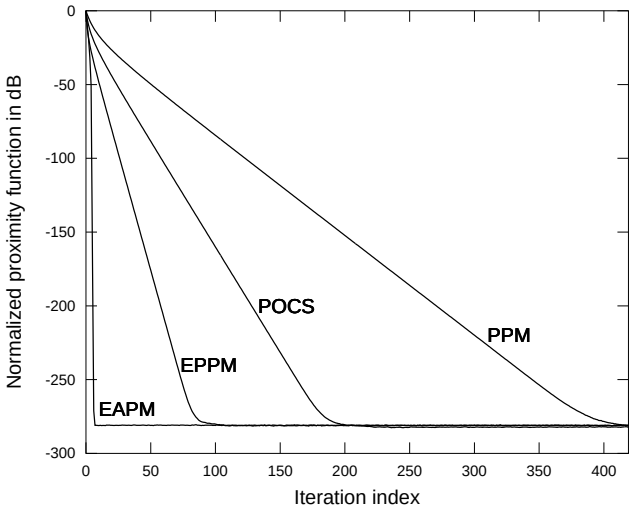


FIG. 1: Average performance of the algorithms when $M \times N = 600 \times 1000$.

strategies that guarantee convergence. In the case of (5), it was shown in [10] that any sequence generated by the *Extrapolated Alternating Projection Method* (EAPM)

$$x^{(0)} \in S_1 \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x^{(n+1)} = x^{(n)} + \rho K_n (P_1 P_2 x^{(n)} - x^{(n)}),$$
$$\text{where} \quad 0 < \rho < 2 \quad \text{and} \quad K_n = \begin{cases} \dfrac{\|P_2 x^{(n)} - x^{(n)}\|^2}{\|P_1 P_2 x^{(n)} - x^{(n)}\|^2}, & \text{if } x^{(n)} \notin S_2, \\ 1, & \text{if } x^{(n)} \in S_2, \end{cases} \tag{9}$$

produces a fast algorithm that converges to a solution to (1). This type of extrapolation scheme, which exploits the fact that S_1 is an affine subspace, actually goes back to the classical work of [50]. It has been further investigated in [11, 35] and has been extended recently to a general block-iterative scheme in [10]. Acceleration methods have also been devised for the parallel algorithm (6). Thus, the convergence of the sequence produced by the *Extrapolated Parallel Projection Method* (EPPM)

$$x^{(0)} \in \mathbb{R}^N \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x^{(n+1)} = x^{(n)} + \chi L_n \left(\frac{P_1 x^{(n)} + P_2 x^{(n)}}{2} - x^{(n)} \right),$$
$$\text{where} \quad 0 < \chi < 2 \quad \text{and}$$
$$L_n = \begin{cases} 2 \dfrac{\|P_1 x^{(n)} - x^{(n)}\|^2 + \|P_2 x^{(n)} - x^{(n)}\|^2}{\|P_1 x^{(n)} + P_2 x^{(n)} - 2x^{(n)}\|^2}, & \text{if } x^{(n)} \notin S_1 \cap S_2, \\ 1, & \text{if } x^{(n)} \in S_1 \cap S_2, \end{cases} \tag{10}$$

to a solution of (1) was established in [34]. This type of parallel extrapolated method goes back to[62] and [66] , and it has been refined or generalized in several places [35, 57, 64]. In particular, it has been shown in numerical experiments to be much faster than unrelaxed projection algorithms in various types of problems ranging from numerical PDEs to image processing [35, 46, 66, 67].

In Figure 1, we compare the numerical performance of POCS (7), PPM (8), EAPM (9), and EPPM (10) for problems of size $M \times N = 600 \times 1000$. As in [10, 35, 33], the performance

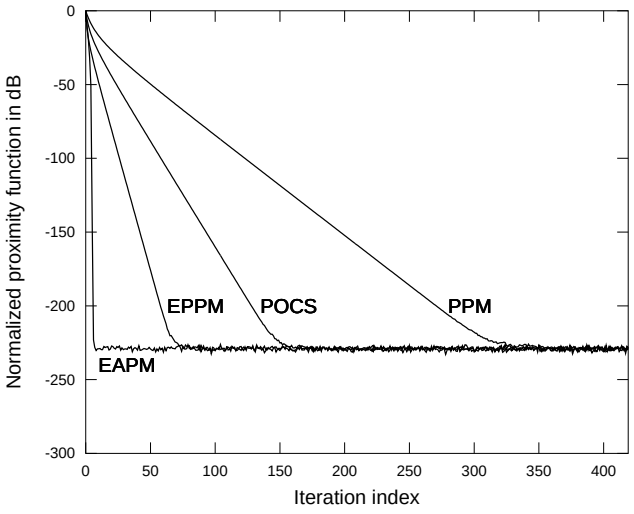


FIG. 2: Average performance of the algorithms when $M \times N = 600 \times 1000$ and condition numbers are around 3×10^4.

of the algorithms is measured by the decibel (dB) values of the normalized proximity function, which is evaluated at the nth iterate $x^{(n)}$ by

$$10 \log_{10} \left(\frac{\|P_1 x^{(n)} - x^{(n)}\|^2 + \|P_2 x^{(n)} - x^{(n)}\|^2}{\|P_1 x^{(0)} - x^{(0)}\|^2 + \|P_2 x^{(0)} - x^{(0)}\|^2} \right). \tag{11}$$

This comparison is relevant because the computational load of each iteration resides essentially in the computation of the projection onto S_1 and it is therefore roughly the same for all four algorithms. The results are averaged over 20 runs of the algorithms initialized with $x^{(0)} = P_1 0$ and $\rho = \chi = 1.9$. In each run a matrix $A \in [-0.5, 0.5]^{M \times N}$ and a vector $x \in [0, 1]^N$ are randomly generated. The vector $b = Ax$ is then constructed so as to obtain a feasible problem using $c_i \equiv 0$ and $d_i \equiv 1$ in (1). As in [48] and many other studies, we observe that POCS is faster than PPM. However, EPPM is faster than POCS and EAPM is clearly the best method : on the average, it is about 60 times faster than PPM, 30 times faster than POCS, and it achieves full convergence in just 7 iterations. In addition, convergence to a feasible solution is guaranteed by the theory and the expression of the extrapolation parameter K_n in (9) is explicit and it requires no additional computation. It is argued in Section 5 of [48] that "there is a significant difference between random and real-life problems (similar observations have been made for linear equations, where random problems tend to be well-conditioned [Reference], and thus often easier to solve than those from applications)." Let us observe that random matrices do show up in many real-life problems, see [37, 76] and the references therein. In addition, as shown in Figure 2, the qualitative behavior of the algorithms in the presence of poor conditioning is quite comparable to that observed in Figure 1 (for the experiments of Figure 2, the condition numbers vary from 3×10^4 to 3.5×10^4).

We have consistently observed this type of performance for problems of various sizes. For instance, we report in Figure 3 on the same experiment as above on problems of size $M \times N = 3000 \times 7000$. Here EAPM is about 45 times faster than PPM, 22 times faster than POCS, and full convergence is achieved in just 5 iterations.

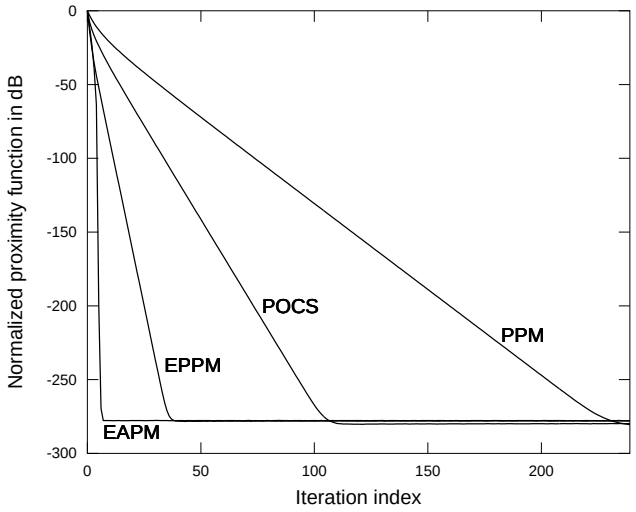


FIG. 3: Average performance of the algorithms when $M \times N = 3000 \times 7000$.

These experiments indicate that the results in [48] on the speed of convergence of POCS and PPM (and the variants proposed there featuring modest speed-up factors and lacking a formal convergence analysis) correspond to a suboptimal implementation of projection methods and are not representative of their performance, since drastic improvements can be achieved by appropriate relaxations.

2.2 An example from image representation

The problem with the largest number of unknowns in the Netlib/CUTEr LP problem set used in [48] has $M \times N = 6,330 \times 22,275$ and (according to the on-line attachment to [48]), for that problem, all methods discussed in [48] need 42 seconds or more to reach the stopping tolerance on a 3.06 GHz Dell Precision 650 workstation. We found among the problems from applications that we have been investigating one that is over an order of magnitude larger and for which the projection algorithm recommended in [28] required only 25 seconds on the average on an Intel Xeon 1.7 GHz processor, 1 Gbyte memory 32 bit workstation using the SNARK09 programming system [41]. We now give a brief description of this problem.

A $J \times J$ digitized image is one that is subdivided into J^2 square-shaped pixels within each of which the image value is uniform. Sometimes alternative representations of an image are superior. For example, in computerized tomography [51], we use the blob basis functions advocated by Lewitt [59] in some series expansion methods to reduce artifacts in the reconstruction. Such a reduction is due to the fact that blob basis functions are smoother than pixel basis functions.

The contribution to the image value at the center of any of the $M = J^2$ pixels by any of the N blob basis functions is known from the geometry of the representations. If we are given a pixel image to start with and would like to find a good blob representation for it, the task is to find the weights x to be given to the blobs so that their combined contributions approximate the pixel values. In mathematical terms, this problem can be formulated as

$$\text{find } \; x \in \mathbb{R}^N \; \text{ such that } \; c \leq Ax \leq d, \tag{12}$$

where the bounds c and d have to be tight to ensure a good approximation of the pixel image by the blob image. (The entries in the matrix A are the values of the various blobs at the centers of the various pixels.)

In the experiments reported in [28] $M \times N = 59{,}049 \times 51{,}152$. The algorithm that was found most efficacious among those tried is the projection method called CART3^{++} : the average (over 40 instances of the problem) time required by CART3^{++} to find a solution to (12) was less than 25 seconds.

The algorithm CART3^{++} belongs to a large family of projection methods that are usually referred to as *algebraic reconstruction techniques* (ART). These were first introduced to the tomographic image reconstruction literature in [47] ; for a recent discussion, see [51, Chapter 11]. CART3^{++}, just like the closely related ART3+ that is used to solve the problem discussed in the next subsection, has an interesting mathematical property : provided that the set of feasible vectors satisfying the inequalities in (12) has a nonempty interior, both CART3^{++}and ART3+ will find a feasible solution in a finite number of iterations [28]. This is achieved by appropriately controlling the sequence of relaxation parameters associated with the individual projections.

2.3 An example from intensity-modulated radiation therapy planning

The goal of intensity-modulated radiation therapy is to deliver sufficient doses to tumors to kill them, but without causing irreparable damage to critical organs. This requirement can be formulated as a linear feasibility problem of the kind shown in (12). The interpretation in this application is that each component of x is a to-be-determined strength of radiation to be delivered to the patient in N separate beamlets, the components of Ax are the resulting doses at M points in the patient's body, and c and d are provided by the radiation oncologist as the desired limits on these doses. Two of the authors of the present paper (W. Chen and G.T. Herman) have been working in this area with D. Craft, T.M. Madden, K. Zhang and H.M. Kooy of the Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, and what follows in this subsection is an outcome of this collaboration.

In the clinical case that we use as an example we have $M \times N = 302{,}491 \times 13{,}734$. The number of nonzero elements in A is 62,226,127, which is less than 1.5% of the total number of entries of A, an important consideration for the efficacy of projection methods for solving the problem. There is an additional technical consideration : since it is impossible to deliver negative radiation, each component of x has to be nonnegative, which results in an additional 13,734 inequality constraints. As mentioned at the end of the last subsection, we use ART3+ [52] to solve this feasibility problem.

In clinical applications, it is considered desirable to find multiple feasible points, each of which is optimal according to its own criterion. A typical optimization task is "find a feasible point that results in the smallest total dose delivered to the liver." The associated functional is a linear one : it is the sum of those components of Ax that are associated with points in the liver. Recognizing the speed by which ART3+ finds a feasible point, we propose to apply it repeatedly, to solve the linear optimization problem

$$\text{find } \ x \in \mathbb{R}^N \ \text{ that minimizes } a^\top x \text{ subject to } c \le Ax \le d. \tag{13}$$

Our method solves this problem by turning the objective function into an additional constraint and solving

$$\text{find } \ x \in \mathbb{R}^N \ \text{ such that } \ c \le Ax \le d \text{ and } a^\top x \le \rho \tag{14}$$

using ART3+. By reducing ρ using a bisection search until we obtain (within a prespecified tolerance) the lowest value possible for it, we get a good approximation to a solution of (13). This whole process is called ART3+O.

The task of minimizing a linear functional subject to linear inequality constraints is the well-known *Linear Programming* (LP) problem and several software packages are available for solving it, see, e.g., [3]. To compare the efficiency of our proposed procedure with currently

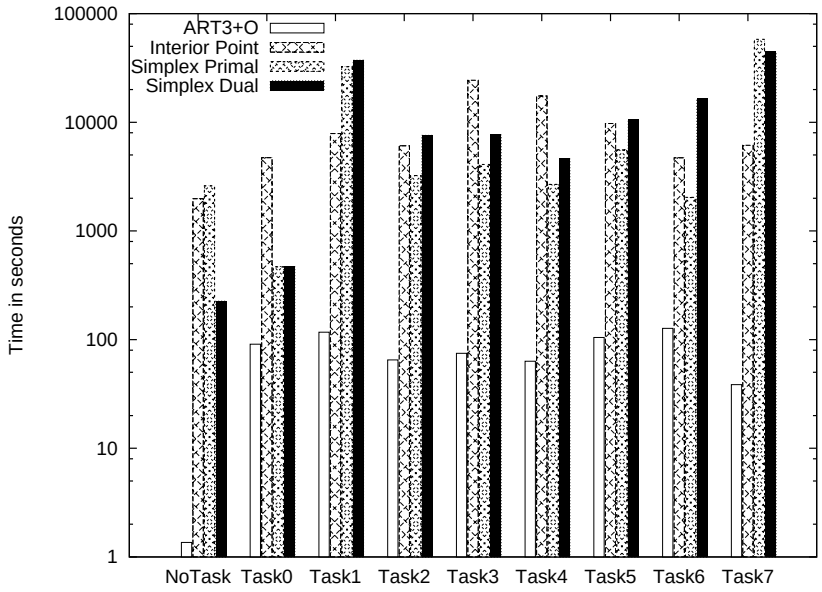


FIG. 4: Timings of the four methods for the feasibility run and the eight optimization tasks.

popular standard approaches, we applied them to the problem (13) for a patient with pancreatic cancer. We used all methods to find just a feasible point (No Task) and also for eight different LP tasks representing various linear optimization criteria. The three algorithms with which we compared ART3+O were the self-dual interior point optimizer, the primal simplex optimizer and the dual simplex optimizer in the commercial software package MOSEK version 5. The results are reported in Figure 4. Typically, for each task, ART3+O used about one to two minutes and the MOSEK algorithms needed one to several hours on an Intel Xeon 2.66 GHz processor, 16 Gbyte memory, 64 bit workstation. It is also noteworthy that the memory requirements of the MOSEK algorithms were at least twelve times as large as that of ART3+O.

2.4 Examples from computerized tomography

Computerized tomography is the problem of recovering an image from its measured (and hence not strictly accurate) integrals along M lines [51]. If we assume that the recovered image will be represented as a linear combination of N basis functions (see Subsection 2.2), then the task is to find the vector x the components of which are the weights to be given to the basis functions. Due to the linearity of integration and based on the knowledge of the basis functions, we can produce an $M \times N$ matrix A such that Ax is approximately the vector b of measurements. Since it is not likely that there is an x such that $Ax = b$, it is reasonable to aim instead at finding an x that minimizes

$$\sigma^2 \|b - Ax\|^2 + \|x\|^2, \tag{15}$$

where $\sigma \in \mathbb{R}$ indicates our confidence in our measurements. As explained in Section 11.3 of [51], this sought-after x is in fact the x part of the minimum norm solution of the consistent system of equations

$$[U \ \sigma A] \begin{bmatrix} u \\ x \end{bmatrix} = \sigma b, \tag{16}$$

where U is the $M \times M$ identity matrix. In the same section there is a derivation of a variant of ART that converges to the sought-after x, given by :

$$\begin{array}{l} u^{(0)} \text{ is the } M\text{-dimensional zero vector,} \\ x^{(0)} \text{ is the } N\text{-dimensional zero vector,} \\ u^{(n+1)} = u^{(n)} + \gamma_n e_{j_n}, \\ x^{(n+1)} = x^{(n)} + \sigma\gamma_n a_{j_n}, \end{array} \tag{17}$$

with

$$\gamma_n = \lambda \frac{\sigma\left(b_{j_n} - a_{j_n}^{\top} x^{(n)}\right) - u_{j_n}^{(n)}}{1 + \sigma^2 \left\|a_{j_n}\right\|^2}, \tag{18}$$

where, for $n \in \mathbb{N}$, $j_n = (n \bmod M) + 1$, for $1 \leq j \leq M$, e_j is the M-dimensional vector whose jth component is 1 and whose other components are 0, $a_j^{\top}$ is the jth row of A and b_j is the jth component of b, and $0 < \lambda < 2$. Recognizing that in one iterative step only one row of the matrix is needed and that in computerized tomography most entries of each row are zero, we see that an iterative step can be carried out very rapidly, provided that we have access to the locations and the values of the nonzero entries. If the memory of the computer is large enough, this can be accommodated by storing A in a row-by-row sparse representation, otherwise the locations and values of the nonzero entries can be generated within each iterative step by some rapid mechanism, such as the *digital difference analyze*r explained, e.g., in Section 4.6 of [51].

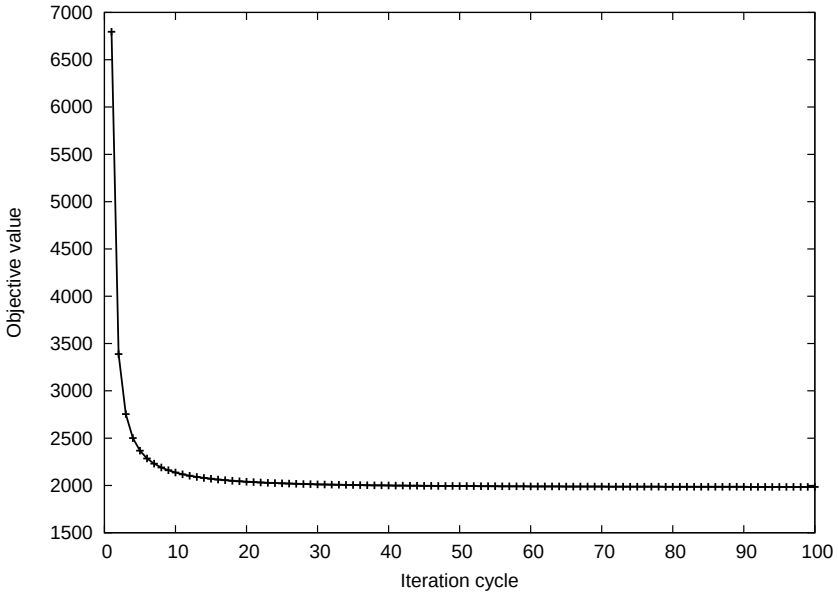
In Section 5.8 of [51] there is an exact specification of the so-called *standard projection data* that are used to evaluate various reconstruction algorithms in that book, the number of lines used in the standard projection data is $M = 223{,}744$. In the evaluations based on the standard projection data that are reported in [51] for reconstruction algorithms that use blob basis functions, the number of blobs used is $N = 51{,}152$. The first experiment on which we report in this subsection used exactly the same arrangement. (For the experiments in this subsection, the input data were created and outputs were analyzed and illustrated using SNARK09 [41].)

In this experiment we applied the ART algorithm of (17) and (18) with $\sigma = 5$ and $\lambda = 0.05$ to the standard projection data. In Figure 5(a) we show the behavior of the objective function (15) as a function of iteration cycles (an *iteration cycle* is defined to be M iterations). It can be observed that the initial decrease in the objective function is very rapid.

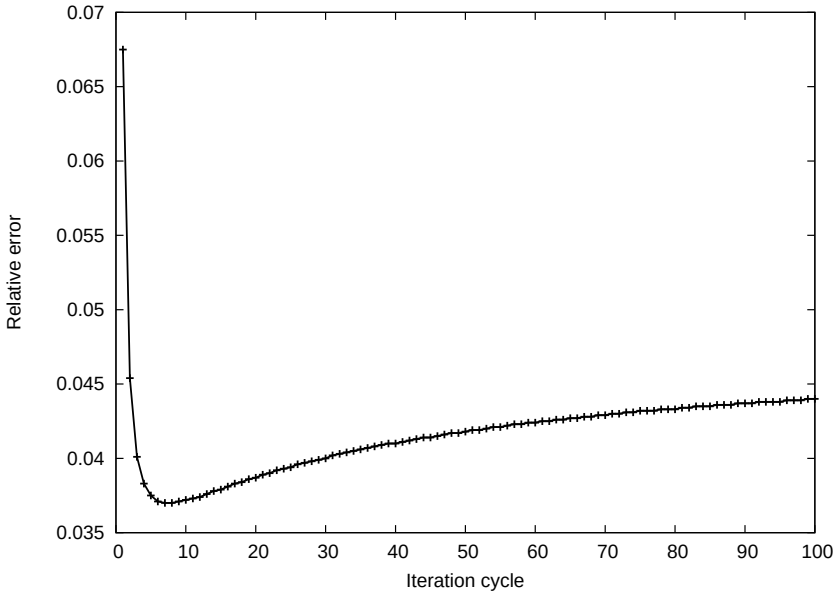
This desirable initial behavior is even more noticeable when we evaluate the algorithm not from the purely mathematical point of view of how well the objective function is reduced, but rather from the application point of view of how good are the reconstructed images. For this purpose, we report on the *normalized mean absolute picture distance measure*, as defined in [51]. To define this measure we need a $J \times J$ digitization of the test phantom for which the data used in the reconstruction were collected ; such a digitization for the phantom we used is shown in Figure 6(a). In our definition of the measure we use $t_{u,v}$ and $s_{u,v}^{(n)}$ to denote the densities of the vth pixel of the uth row of the digitized test phantom and of the reconstruction (which is obtained from the vector $x^{(n)}$ of blob coefficients), respectively. We define the distance measure as

$$r^{(n)} = \frac{\sum_{u=1}^{J} \sum_{v=1}^{J} \left| t_{u,v} - s_{u,v}^{(n)} \right|}{\sum_{u=1}^{J} \sum_{v=1}^{J} |t_{u,v}|}. \tag{19}$$

In Figure 5(b) we plot $r^{(n)}$ for this experiment. It is seen that its minimum is reached at the seventh iteration cycle, i.e., when $n = 7M$. This reflects the fact that the minimization objective (15) does not (and, in fact, it cannot in real applications where the phantom is not known to us) fully describe the application objective. For this reason it is standard practice

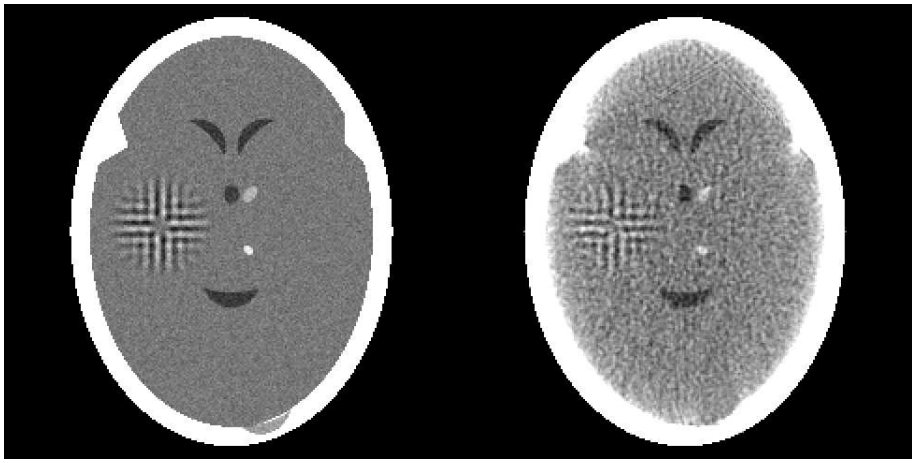


(a) Plot of the objective function (15).



(b) Plot of the distance measure (19).

FIG. 5: Image reconstruction by ART when $M \times N = 223{,}744 \times 51{,}152$.



(a) (b)

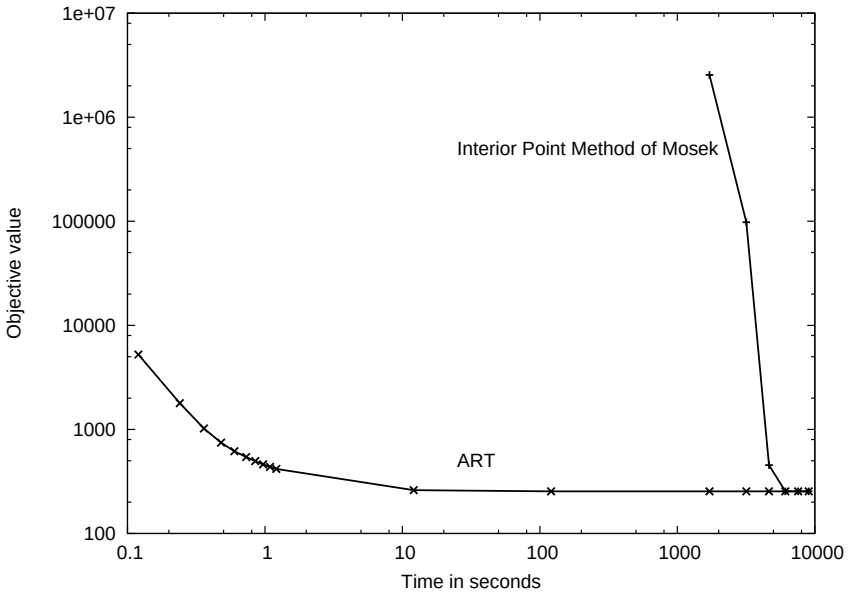
FIG. 6: Displays of a 243×243 digitized phantom (a) and of an ART reconstruction when $M \times N = 223{,}744 \times 51{,}152$ (b).

in tomography [51] to stop the iterative process after a few iteration cycles and use the result at that time as the reconstruction. The digitization obtained from $x^{(7M)}$ produced by this experiment is shown in Figure 6(b). The reconstruction is not perfect (as indeed it cannot possibly be since the measured data are only approximations of the line integrals assumed by the mathematics), but important features of the phantom are identifiable in the reconstruction. This ART reconstruction was carried out in 38.4 seconds on an Intel Core 1.6 GHz processor, 2 Gbyte memory, 32 bit laptop.

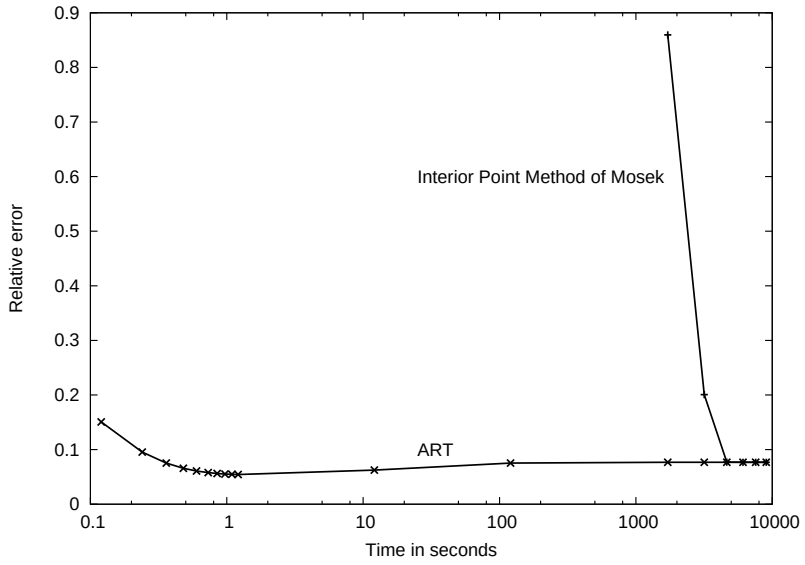
We wanted to compare the time needed by ART with the time needed to solve the system (16) of consistent equations for the same data by the current implementation of the interior point method of MOSEK version 5 [3]. Unfortunately this could not be done, because the memory requirements of the MOSEK software were too large for our laptop. So we attempted to use a much more powerful Intel Xeon 2.66 GHz processor, 16 Gbyte memory, 64 bit workstation, but even the 16 Gbyte memory was too small to handle this problem using the MOSEK software. The importance of this memory requirement issue for the subject matter of this paper cannot be overemphasized : problems that routinely arise in real applications can be handled by projection methods using inexpensive laptops, while "more sophisticated alternatives" fail to produce any results even on much more powerful workstations due to their much greater demands on computer memory.

In order to be able to compare the efficiency of ART with that of the interior point method in MOSEK we had to reduce M and N to about a ninth of their previously-used sizes. Thus, in the second experiment on which we now report $M \times N = 24{,}880 \times 5{,}711$. For this smaller example we ran both ART and the interior point method in MOSEK (with its default parameters) on the Intel Xeon 2.66 GHz processor, 16 Gbyte memory, 64 bit workstation. In Figure 7 we plot both the objective function and the distance measure for both algorithms as a function of time. From the point of view of the objective function, MOSEK needed over 5000 seconds to reach a value as low as ART reached in 10 seconds. The advantage of ART is more pronounced when considering the picture distance measure : the optimal value is reached by ART at 1.7 seconds (when $n = 14M$) while the interior point method never reaches a distance value that is as low as that of ART and it needs approximately 5000 seconds to reach its lowest distance measure.

Since both M and N are about a ninth of their previous sizes, we report in Figure 8 on

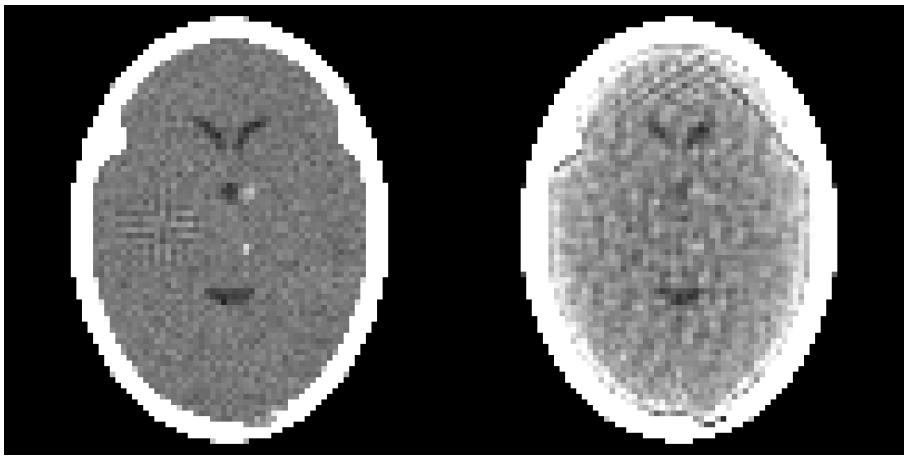


(a) Plot of the objective function (15).



(b) Plot of the distance measure (19)

FIG. 7: Image reconstruction by ART and the interior point method in MOSEK when $M \times N = 24{,}880 \times 5{,}711$.



(a) (b)

FIG. 8: Displays of an 81×81 digitized phantom (a) and of an ART reconstruction when $M \times N = 24{,}880 \times 5{,}711$ (b).

the 81×81 digitizations of the phantom and of the reconstruction $x^{(14M)}$. These are clearly inferior to the images in Figure 6, demonstrating the medical necessity for the larger system of equations.

3 Published and patented results

3.1 Scientific publications

Here we give a brief glimpse into some recently published results that show the efficacy of projection methods for some large problems. In the problems discussed in [40], the number of unknowns was 59,049. In the examples given in [52] (a paper devoted to radiation therapy planning), problems of the form (12) were considered with the number N of unknowns only 515 but the number of pairs of constraints $M = 128{,}688$. In four out of the six cases reported there, the projection method ART3+ [52]found a feasible point in less than three seconds, and in the remaining two cases a feasible point was found in less than 34 seconds. These times are for a standard PC, using an Intel Xeon 1.7 GHz processor and 1 Gbyte memory. The problems in [40, 52] are small compared to some of the other applications for which projection methods have been successfully used. In [19] (a paper devoted to reconstruction from electron micrographs), there are examples in which 16,777,216 unknowns are to be recovered from 4,587,520 measurements (each giving an approximate linear equality) and others in which 884,436 unknowns are to be recovered from 92,160,000 measurements. Projection methods were used in [19] to handle such large problems in a reasonable time.

In a recent paper [56] it is shown that a variant of ART can be used for crystal lattice orientation distribution function estimation from diffraction data. One of the problems discussed in [56] has 1,372,000,000 unknowns and the number of equations is potentially infinite. They are randomly generated and a projection step can be carried out as soon as a new equation is available (an ideal use of a sequential projection method of the row-action type, see [20]). The result reported in the paper for that problem is the one obtained after 1,000,000,000 such projection steps.

As for all methodologies, projection methods are not necessarily the approach of choice

in all applications. However, in important applications in biomedicine and image processing, projection methods work well and have been used successfully for a long time. For example, an important application of reconstruction from projections is electron microscopy and some of the leading groups in that field consider the projection method "ART with blobs" to be the method of choice, see [13]. A mathematical reason for this is that for such problems the angles between hyperplanes or half-spaces, represented by linear equalities or linear inequalities as in (1) and (12), are in general large (in the sense that the cosine of the angle between the normals of two randomly chosen hyperplanes in the system to be solved is likely to be near zero) due to the high sparsity in each of the rows of the system matrix.

3.2 Commercial patents

There is hardly better evidence for the value of projection methods than the many patents for commercial purposes that include them. Projection methods are used in commercial devices in many areas. Unfortunately, if a device is truly commercial, then the algorithm that is actually used in it is proprietary and usually not published. Many commercial emission tomography scanners use now some sort of iterative algorithms. A prime example is provided by the commercially-successful Philips Allegro scanners (see http ://www.healthcare.philips.com/main/products/ and [29]). In x-ray computerized tomography (CT), there are reports emanating from companies that sell such scanners indicating that variants of ART are used in heart imaging ; an example is presented in [54].

The first EMI (Electric & Musical Industries Ltd., London, England, UK) CT scanner, invented by G.N. Hounsfield [53], used a variant of ART. For this pioneering invention, Hounsfield shared the Nobel Prize with A.M. Cormack in 1979. Thirty years later (on September 29, 2009), a patent was issued to Philips (Koninklijke Philips Electronics N.V., Eindhoven, The Netherlands) for a "Method and device for the iterative reconstruction of cardiac images" [77]. The role of projection methods is demonstrated by the following quote from the "Summary of the Invention" included in the Patent Description :

> "The iterative reconstruction applied here may particularly be based on an Algebraic Reconstruction Technique (ART) (cf. R. Gordon, R. Bender, and G.T. Herman : "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography", J. Theor. Biol., 29 :471–481, 1970) or on a Maximum Likelihood (ML) algorithm (K. Lange and J.A. Fessler : "Globally convergent algorithms for maximum a posteriori transmission tomography", IEEE Transactions on Image Processing, 4(10) :1430–1450, 1995), wherein each image update step uses the projections of a selected subset, i.e., projections corresponding to a similar movement phase."

4 Conclusion

In this paper we have shown that, whether or not alternative methods are applicable, correctly implemented projection methods are very efficient for convex feasibility problems with linear inequality constraints, especially for those that are large, sparse, and originate from real-life applications.

Acknowledgments. The work of Y. Censor, W. Chen, R. Davidi, and G.T. Herman is supported by NIH Award Number R01HL070472 from the National Heart, Lung, and Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health. The work of P.L. Combettes was supported by the Agence Nationale de

la Recherche under grant ANR-08-BLAN-0294-02. We thank the Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School (and especially Dr. David Craft) for providing access to their computers.

Références

[1] Agmon, S. : The relaxation method for linear inequalities. Canad. J. Math. **6**, 382–392 (1954)

[2] Aharoni, R., Censor, Y. : Block-iterative projection methods for parallel computation of solutions to convex feasibility problems. Linear Algebra Appl. **120**, 165–175 (1989)

[3] Andersen, E.D., Andersen, K.D. : The MOSEK interior point optimizer for linear programming : an implementation of the homogeneous algorithm. In : Frenk, H., Roos, K., Terlaky, T., Zhang, S. (eds.) High Performance Optimization, pp. 197–-232, Kluwer, Boston (2000).

[4] Andrews, H.C., Hunt, B.R. : Digital Image Restoration. Prentice-Hall, Englewood Cliffs (1977)

[5] Auslender, A. : Méthodes Numériques pour la Résolution des Problèmes d'Optimisation avec Contraintes. Thèse, Faculté des Sciences, Grenoble (1969)

[6] Auslender, A. : Optimisation - Méthodes Numériques. Masson, Paris (1976)

[7] Bauschke, H.H. : The approximation of fixed points of compositions of nonexpansive mappings in Hilbert space. J. Math. Anal. Appl. **202**, 150–159 (1996).

[8] Bauschke, H.H., Borwein, J.M. : On projection algorithms for solving convex feasibility problems. SIAM Rev. **38**, 367–426 (1996)

[9] Bauschke, H.H., Combettes, P.L. : A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces. Math. Oper. Res. **26**, 248–264 (2001)

[10] Bauschke, H.H., Combettes, P.L., Kruk, S.G. : Extrapolation algorithm for affine-convex feasibility problems. Numer. Algorithms **41**, 239–274 (2006)

[11] Bauschke, H.H., Deutsch, F., Hundal, H., Park, S.-H. : Accelerating the convergence of the method of alternating projections. Trans. Amer. Math. Soc. **355**, 3433–3461 (2003)

[12] Bauschke, H.H., Matou¨ková, E., Reich, S. : Projection and proximal point methods : convergence results and counterexamples. Nonlinear Anal. **56**, 715–738 (2004)

[13] Bilbao-Castro, J.R., Marabini, R., Sorzano, C.O.S., García, I., Carazo, J.M., Fernández, J.J. : Exploiting desktop supercomputing for three-dimensional electron microscopy reconstructions using ART with blobs. J. Struct. Biol. **165**, 19–26 (2009)

[14] Blatt, D., Hero, A.O., III : Energy based sensor network source localization via projection onto convex sets (POCS). IEEE Trans. Signal Process. **54**, 3614–3619 (2006)

[15] Brègman, L.M. : The method of successive projection for finding a common point of convex sets. Soviet Math. Dokl. **6**, 688–692 (1965)

[16] Butnariu, D., Censor, Y. : On the behavior of a block-iterative projection method for solving convex feasibility problems. Int. J. Comput. Math. **34**, 79–94 (1990)

[17] Butnariu, D., Censor, Y., Reich, S. (eds.) : Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications. Elsevier, Amsterdam (2001)

[18] Butnariu, D., Davidi, R., Herman, G.T., Kazansev, I.G. : Stable convergence behavior under summable perturbations of a class of projection methods for convex feasibility and optimization problems, IEEE J. Selected Topics Signal Process. **1**, 540–547 (2007)

[19] Carazo, J.M., Herman, G.T., Sorzano, C.O.S., Marabini R. : Algorithms for three-dimensional reconstruction from imperfect projection data provided by electron microscopy. In : Frank, J. (ed.) Electron Tomography : Methods for Three-Dimensional Visualization of Structures in the Cell, 2nd edition, pp. 217–243, Springer Science+Business Media, LLC, New York (2006)

[20] Censor, Y. : Row-action methods for huge and sparse systems and their applications. SIAM Rev. **23**, 444–466 (1981)

[21] Censor, Y., Altschuler, M.D., Powlis, W.D. : On the use of Cimmino's simultaneous projections method for computing a solution of the inverse problem in radiation therapy treatment planning. Inverse Problems **4**, 607–623 (1988)

[22] Censor, Y., Elfving, T., Herman, G.T. : Averaging strings of sequential iterations for convex feasibility problems. In : Butnariu, D., Censor, Y., Reich, S. (eds.) Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications, pp. 101–114, Elsevier Science Publishers, Amsterdam (2001)

[23] Censor, Y., Gordon, D., Gordon, R. : BICAV : A block-iterative, parallel algorithm for sparse systems with pixel-related weighting. IEEE Trans. Med. Imaging **20**, 1050–1060 (2001)

[24] Censor, Y., Segal, A. : On the string averaging method for sparse common fixed points problems. Int. Trans. Oper. Res. **16**, 481–494 (2009)

[25] Censor, Y., Tom, E. : Convergence of string-averaging projection schemes for inconsistent convex feasibility problems. Optim. Methods Softw. **18**, 543–554 (2003)

[26] Censor, Y., Zenios, S.A. : Parallel Optimization : Theory, Algorithms, and Applications. Oxford University Press, New York (1997)

[27] Cetin, A.E., Ozaktas, H., Ozaktas, H.M. : Resolution enhancement of low resolution wavefields with POCS algorithm. Electron. Lett. **39**, 1808–1810 (2003)

[28] Chen, W., Herman, G.T. : Effcient controls for finitely convergent sequential algorithms. ACM Trans. Math. Software, to appear

[29] Chiang, S., Cardi, C., Matej, S., Zhuang, H., Newberg, A., Alavi, A., Karp, J.S. : Clinical validation of fully 3-D versus 2.5-D RAMLA reconstruction on the Phillips-ADAC CPET PET scanner. Nucl. Med. Commun. **25**, 1103–1107 (2004)

[30] Choi, H., Baraniuk, R.G. : Multiple wavelet basis image denoising using Besov ball projections. IEEE Signal Process. Lett. **11**, 717–720 (2004)

[31] Cimmino, G. : Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari. La Ricerca Scientifica (Roma) **1**, 326–333 (1938)

[32] Combettes, P.L. : The foundations of set theoretic estimation. Proc. IEEE **81**, 182–208 (1993)

[33] Combettes, P.L. : The convex feasibility problem in image recovery. Adv. Imaging Electron. Phys. **95**, 155–270 (1996)

[34] Combettes, P.L. : Hilbertian convex feasibility problem : Convergence of projection methods. Appl. Math. Optim. **35**, 311–330 (1997)

[35] Combettes, P.L. : Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections. IEEE Trans. Image Process. **6**, 493–506 (1997)

[36] Combettes, P.L., Hirstoaga, S.A. : Equilibrium programming in Hilbert spaces. J. Nonlinear Convex Anal. **6**, 117–136 (2005)

[37] Combettes, P.L., Trussell, H.J. : Methods for digital restoration of signals degraded by a stochastic impulse response. IEEE Trans. Acoust. Speech Signal Process. **37**, 393–401 (1989)

[38] Cottle, R.W., Pang, J.-S. : On solving linear complementarity problems as linear programs. Math. Program. Study **7**, 88–107 (1978)

[39] Crombez, G. : Finding common fixed points of strict paracontractions by averaging strings of sequential iterations, J. Nonlinear Convex Anal. **3,** 345–351 (2002)

[40] Davidi, R., Herman, G.T., Censor, Y. : Perturbation-resilient block-iterative projection methods with application to image reconstruction from projections. Int. Trans. Oper. Res. **16**, 505–524 (2009)

[41] Davidi, R., Herman, G.T., Klukowska, J. : SNARK09 : A programming system for the reconstruction of 2D images from 1D projections. The CUNY Institute for Software Design and Development. http ://www.snark09.com (2009). Accessed 13 December 2009

[42] Deutsch, F. : Best Approximation in Inner Product Spaces. Springer-Verlag, New York (2001)

[43] Eckstein, J., Svaiter, B. F. : General projective splitting methods for sums of maximal monotone operators. SIAM J. Control Optim. **48**, 787–811 (2009)

[44] Eggermont, P.P.B., Herman, G.T., Lent, A. : Iterative algorithms for large partitioned linear systems, with applications to image reconstruction, Linear Algebra Appl. **40**, 37–67, (1981)

[45] Golub, G.H., van Loan, C.F. : Matrix Computations, 3rd ed. Johns Hopkins University Press, Baltimore, MD (1996)

[46] González-Castaño, F.J., García-Palomares, U.M., Alba-Castro, J.L., Pousada-Carballo, J.M. : Fast image recovery using dynamic load balancing in parallel architectures, by means of incomplete projections. IEEE Trans. Image Process. **10**, 493–499 (2001)

[47] Gordon, R., Bender, R., Herman, G.T. : Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography. J. Theor. Biol. **29**, 471–482 (1970)

[48] Gould, N.I.M. : How good are projection methods for convex feasibility problems ? Comput. Optim. Appl. **40**, 1–12 (2008)

[49] Gu, J., Stark, H., Yang, Y. : Wide-band smart antenna design using vector space projection methods. IEEE Trans. Antennas Propag. **52**, 3228–3236 (2004)

[50] Gubin, L.G., Polyak, B.T., Raik, E.V. : The method of projections for finding the common point of convex sets. USSR Comput. Math. Math. Phys. **7**, 1--24 (1967)

[51] Herman, G.T. : Fundamentals of Computerized Tomography : Image Reconstruction from Projections, 2nd ed. Springer, New York (2009)

[52] Herman, G.T., Chen, W. : A fast algorithm for solving a linear feasibility problem with application to intensity-modulated radiation therapy. Linear Algebra Appl. **428**, 1207–1217 (2008)

[53] Hounsfield, G.N. : A method and apparatus for examination of a body by radiation such as X or gamma radiation. UK Patent No. 1283915 (1968/72)

[54] Isola, A.A., Ziegler, A., Koehler, T., Niessen, W.J., Grass, M. : Motion compensated iterative cone-beam CT image reconstruction with adapted blobs as basis functions. Phys. Med. Biol. **53**, 6777–6797 (2008)

[55] Kaczmarz, S. : Angenäherte Auflösung von Systemen linearer Gleichungen. Bull. Acad. Sci. Pologne **A35**, 355–357 (1937)

[56] Kazantsev, I.G., Schmidt, S., Poulsen, H.F. : A discrete spherical x-ray transform of orientation distribution functions using bounding cubes. Inverse Problems **25**, 105009 (2009)

[57] Kiwiel, K.C., Łopuch, B. : Surrogate projection methods for finding fixed points of firmly nonexpansive mappings. SIAM J. Optim. **7**, 1084–1102 (1997)

[58] Lee, S.-H., Kwon, K.-R. : Mesh watermarking based on projection onto two convex sets. Multimedia Syst. **13**, 323–330 (2008)

[59] Lewitt, R.M. : Multidimensional digital image representation using generalized Kaiser-Bessel window functions. J. Opt. Soc. Amer. A **7**, 1834–1846 (1990)

[60] Liew, A.W.-C., Yan, H., Law, N.-F. : POCS-based blocking artifacts suppression using a smoothness constraint set with explicit region modeling. IEEE Trans. Circuits Syst. Video Technol. **15**, 795–800 (2005)

[61] Lu, Y.M., Karzand, M., Vetterli, M. : Demosaicking by alternating projections : Theory and fast one-step implementation. IEEE Trans. Image Process., to appear

[62] Merzlyakov, Y.I. : On a relaxation method of solving systems of linear inequalities. USSR Comput. Math. Math. Phys. **2**, 504–510 (1963)

[63] Motzkin, T.S., Schoenberg, I.J. : The relaxation method for linear inequalities. Canad. J. Math. **6**, 393–404 (1954)

[64] Ottavy, N. : Strong convergence of projection-like methods in Hilbert spaces. J. Optim. Theory Appl. **56**, 433–461 (1988)

[65] Penfold, S.N., Schulte, R.W., Censor, Y., Bashkirov, V., McAllister, S., Schubert, K.E., Rosenfeld, A.B. : Block-iterative and string-averaging projection algorithms in proton computed tomography image reconstruction. In : Censor, Y., Jiang, M., Wang, G. (eds.) Biomedical Mathematics : Promising Directions in Imaging, Therapy Planning and Inverse Problems, Medical Physics Publishing, Madison, WI, to appear.

[66] Pierra, G. : Éclatement de contraintes en parallèle pour la minimisation d'une forme quadratique. In : Proc. 7th IFIP Conf. on Optimization Techniques : Modeling and Optimization in the Service of Man, Lecture Notes in Comput. Sci. **41**, pp. 200–218. Springer-Verlag, London (1976)

[67] Pierra, G. : Decomposition through formalization in a product space. Math. Programming **28**, 96–115 (1984)

[68] Rhee, H. : An application of the string averaging method to one-sided best simultaneous approximation. J. Korea Soc. Math. Educ. Ser. B : Pure Appl. Math. **10**, 49–56 (2003)

[69] Samsonov, A.A., Kholmovski, E.G., Parker, D.L., Johnson, C.R. : POCSENSE : POCS-based reconstruction for sensitivity encoded magnetic resonance imaging. Magn. Reson. Med. **52**, 1397–1406 (2004)

[70] Shaked, N.T., Rosen, J. : Multiple-viewpoint projection holograms synthesized by spatially incoherent correlation with broadband functions. J. Opt. Soc. Amer. A **25**, 2129–2138 (2008)

[71] Sharma, G. : Set theoretic estimation for problems in subtractive color. Color Res. Appl. **25**, 333-348 (2000)

[72] Stark, H., Yang, Y. : Vector Space Projections : A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics. Wiley-Interscience, New York (1998)

[73] van Wyk, B.J., van Wyk, M.A. : A POCS-based graph matching algorithm. IEEE Trans. Pattern Anal. Machine Intell. **26**, 1526–1530 (2004)

[74] Youla, D.C., Webb, H. : Image restoration by the method of convex projections : Part 1 – theory. IEEE Trans. Med. Imaging **1**, 81–94 (1982)

[75] Yukawa, M., Yamada, I. : Pairwise optimal weight realization – Acceleration technique for set-theoretic adaptive parallel subgradient projection algorithm. IEEE Trans. Signal Process. **54**, 4557–4571 (2006)

[76] Zhang, T., Hong, H. : Restoration algorithms for turbulence-degraded images based on optimized estimation of discrete values of overall point spread functions. Opt. Eng. **44**, 017005 (2005)

[77] Ziegler, A., Grass, M., Koehler, T. : Method and device for the iterative reconstruction of cardiac images. US Patent Number 7596204 (2009)